2016年度 CYBEX Working Group 活動報告

伊藤 俊一郎 (ito.shunichiro.in0@is.naist.jp) 宮本 大輔 (daisu-mi@nc.u-tokyo.ac.jp)
門林 雄基 (youki-k@is.aist-nara.ac.jp)

2016年12月15日

目 次

1 CYBEX WG 2016年度の活動

近年，様々なサイバーセキュリティ技術が開発されており，その普及は急務である．これまでの CYBEX WG は，様々なサイバーセキュリティ技術の先端的研究成果や運用によって得られた知見について，国際標準化活動などの様々なチャネルを通じ，多くのステークホルダーに啓蒙することを目的とした活動に従事してきた．

我々の研究グループでは，サイバー脅威情報の共有に取り組んできた．サイバー攻撃は多くの組織を狙って攻撃が行われており，組織連携なくしてインシデント対応は行えず，情報共有なくして組織共有は行えない．この一方で，STIX や IODEF に見られるように，M2M (Machine to Machine) における情報の連携の取り組みは活発であるが，M2H (Machine to Human) において情報を効率よく伝達する方法もまた取り組むべきであると考える．この観点から，我々は以下の活動を行った．

今年度の主な活動内容は以下の通りである．

- ITU-T SG17 X.Cogent の標準化提案
- なりすましメール対策におけるセキュリティインディケータの有効性の評価

2 ITU-T SG17 X.Cogent の標準化提案

サイバー犯罪に用いられる技術，それに対抗するサイバーセキュリティ技術は著しく進歩している．その一方で，エンドユーザのセキュリティ技術への理解は追いついていない現状が指摘されている．そこで，ITU-T Study Group 17 Question 4 において，セキュリティ技術をエンドユーザに提示する画像や文章，色といった要素技術での工夫や，アクセシビリティについて調査を行い，X.Cogent という標準化提案を行っている．2016年度もこの提案の改訂作業を行い，ITU-T X.1212 国際標準としての成立を目指している．

なお，報告書執筆時点における国際標準化提案内容を報告書の付録として末尾に掲載する．

3 なりすましメール対策におけるセキュリティインディケータの有効性の評価

なりすましメールとは悪意あるメールユーザが特定の人物や組織になりすまして送信する悪性メールのこと

図 1: DKIM Verifier における認証結果の表示

図 2: ThunderSec における認証結果の表示

である．なりすましメールは巧妙化してきており，メール受信者が文面等から正規メールとなりすましメールを区別するのが困難な状況である．既存対策ではメール送信者の認証を行い，その認証結果を受信メールに表示することで，メール受信者によるなりすましメールの判別を支援している．しかし，認証情報をメール受信者が知覚できているか，認証結果の意味を正確に認識できているのかは充分に検討されていない．

本研究では，メールの安全性を示す指標をセキュリティインディケータと呼称し，なりすましメール対策に資するセキュリティインディケータを提案する．また，既存対策を含めたセキュリティインディケータ（以下,「インディケータ」という.）の有効性を調査する．

3.1 既存対策と提案手法

なりすましメール対策としては主に送信ドメイン認証や電子署名がある．これらの対策はメール送信者の身元を検証し，その認証結果をメール受信者に示している．送信ドメイン認証の結果は受信メールのヘッダに表示される．しかし，すべてのユーザがメールヘッダを確認しているわけではなく，確認するにしても労力がかかる．これを補う方策として，メーラーである Thunderbird には送信ドメイン認証結果を可視化するアドオンが存在する．図 1 は DKIM Verifier における送信ドメイン認証結果の可視化である．メール差出人の背景色を緑色で表示することで，送信ドメイン認証が成功したことを示している．図 2 は ThunderSec における送信ドメイン認証結果等の可視化である．赤い通知バーを表示することで，送信ドメイン認証等が失敗したことを示している．また，電子署名を用いた認証手法の一つとして S/MIME がある．Thunderbird においては，S/MIME の認証結果は署名マークのアイコンで表示される．

以上のように，既存のなりすましメール対策においては，メール受信者によるなりすましメールの判別に資するインディケータが表示されている．しかし，既存対策のインディケータの有効性は充分に評価されていないため，なりすましメール対策に資するインディケータについて検討し，新たなインディケータを提案するとともに既存対策を含めた有効性の評価を行う．

なりすましメールを判別するためのインディケータとして重要な要素は 3 つある．1 つ目は,「存在の認識」である．メール送信者の認証結果を示すインディケータがメール上に表示されていることをメール受信者に知覚させることを意味する．メール受信者がインディケータを知覚することができない場合は，受信メールが認証されていないメールと同義で扱われるというエラーが発生する．2 つ目は,「インディケータが持つ意味の正確な伝達」である．インディケータには受信メールの安全性を受信者に伝える役割があるが，インディケータが持つ意味を正確にメール受信者に伝えることができなければ，重大なエラーが発生する．例えば，受信メールが危険であることを示すインディケータが表示された時，メール受信者が安全なメールであると誤認識してしまうエラーが発生してしまった場合は，インディケータの役割を果たせなくなる．よって，インディケータは受信メールが安全なのか，危険なのかを明確に示すことが重要となる．最後に「認知的負荷の軽減」である．メール受信者に労力をかけることなくインディケータが持つ意味を認識させることを意味する．例として，メール送信者の認証が失敗した場合に，その危険性を伝えるために専門用語を含んだ長文の警告文が出たとする．メール受信者は文章を読まなくてはならず，その意味を理解するために用語の意味を調べるという労力が発生する．このように，インディケータの持つ意味を認識する際に認知的負荷がかかると，メール受信者によってインディケータが使用されなくなる，もしくは警告を無視する可能性が高い．以上より，インディケータはメール受信者の認知的負荷を局限するデザインであることが重要と言える．

以上の検討を踏まえて，提案するインディケータでは画像を利用する．表示位置や大きさを設定できる画像は，メール受信者にその存在を知覚させやすいとい

図 3: 提案するセキュリティインディケータ

う特徴がある．また，安全や危険をそれぞれ連想させる画像を用いることで，メール受信者にメールの安全性を伝えることもでき，かつ認知的負荷の軽減も期待できる．提案するインディケータを図 3 に示す．左側の緑十字の画像は日本においては安全を意味する印であり，右側は危険性を示すドクロマークと黄色を組み合わせた画像である．メール送信者の認証が成功した場合には左側の画像を，認証が失敗した場合には右側の画像をメールの差出人欄付近に表示することで，メール受信者に受信メールの安全性を伝える．

3.2 セキュリティインディケータの有効性の調査及び評価

提案手法と既存対策のインディケータの有効性を調査するために，Web アンケートを実施した．アンケートの概要は以下のとおりである．

- アンケート対象者
  WIDE Project 2016 年 12 月研究会 SWAN BoF 参加者及び所属研究室学生
- 有効回答数
  31
- 調査対象のインディケータ
  - 提案手法
    安全なメールを表すインディケータ: SI1
    危険なメールを表すインディケータ: SI5
  - 署名マーク（S/MIME で使用）
    安全なメールを表すインディケータ: SI6
    危険なメールを表すインディケータ: SI2
  - 差出人の背景色強調（DKIM Verifier で使用）
    安全なメールを表すインディケータ: SI4
    危険なメールを表すインディケータ: SI8
  - 通知バー（ThunderSec で使用）
    安全なメールを表すインディケータ: SI7
    危険なメールを表すインディケータ: SI3
- 調査事項
  - インディケータの視認性及びインディケータが持つ意味の正確な伝達
  - ユーザ特性（メール確認環境下でのストレスの多寡）

調査対象となるそれぞれのインディケータはメール送信者の認証結果に応じ，安全なメール，または危険なメールという 2 種類の結果を表示するものとする．調査では，8 つの Thunderbird のメール画面をアンケート協力者に提示し，SI（セキュリティインディケータ）に関する質問を行った．8 つの画面には安全もしくは危険を示す SI1~8 が一つずつ表示されている．また，メール確認環境下でのストレスの多寡というユーザ特性についても調査した．なりすましメールは人のエラーを誘引することでマルウェア感染を引き起こす．よって，エラーを誘引する要因とも言えるストレスのメール確認時における多寡によって，効果的なインディケータがユーザごとに異なるのかを調査した．ストレスの多寡については，1 日のメール処理数，メール確認時の時間的制約，メール確認時の疲労度の 3 問で評価した．各質問に対してよりストレスが多い回答を 4 点，よりストレスが少ない回答を 1 点として 4 段階で点数付けし，3 問の平均点で評価した．

アンケート調査結果の一部を表 1 に示す．効果的なインディケータの定義としては，インディケータに関する質問に対して，視認性が高いという回答をし，かつインディケータが持つ意味（安全もしくは危険）を正確に回答できた場合に効果的なインディケータとする．

安全なメールを表すインディケータについては提案手法である SI1 が，危険なメールを表すインディケータについては通知バーである SI3 と提案手法である SI5 が最も効果的であるという結果が得られた．このことから，メールの安全性を受信者に伝える上で，画像を使用したインディケータが既存対策に比べて効果があるということが分かった．SI3（図 2）が効果的であるのは，赤い通知バーを表示するため視認性が良く，警告文も表示されるため，受信者がインディケータの意味を理解するのに役立ったためであると考える．一方

表 1: 効果的なセキュリティインディケータに関する Web アンケートの調査結果

	全体 n=31			ストレス多い（> 2.5） n=23			ストレス少ない（≤ 2.5） n=8		
安全	SI の種類	回答者数	割合	SI の種類	回答者数	割合	SI の種類	回答者数	割合
	SI1	4	12.9%	SI1	3	13.0%	SI1	1	12.5%
	SI4	2	6.5%	SI4	2	8.7%	SI7	1	12.5%
	SI6	1	3.2%	SI6	1	4.4%	SI4	0	0.0%
	SI7	1	3.2%	SI7	0	0.0%	SI6	0	0.0%
危険	SI の種類	回答者数	割合	SI の種類	回答者数	割合	SI の種類	回答者数	割合
	SI3	17	54.8%	SI3	13	56.5%	SI5	6	75.0%
	SI5	17	54.8%	SI5	11	47.8%	SI3	4	50.0%
	SI8	3	9.7%	SI8	2	8.7%	SI8	1	12.5%
	SI2	0	0.0%	SI2	0	0.0%	SI2	0	0.0%

で，安全を示す通知バーである SI7 については，効果的でないインディケータであるという結果が得られた．SI7 は図 2 の赤い通知バーが黒色に変わり，受信メールが安全であることを示す文章が記述されている．SI7 に対する回答からは，SI7 はメールの安全性を伝えることはできていることが分かった．しかし，同じ通知バーである SI3 では，認識しづらい傾向があるという回答をしたユーザは約 3%であったのに対し，SI7 は約 36%であった．このことから，通知バーの文章によりインディケータの意味は伝えられているものの，SI7 の黒い通知バーは認識しづらいため，インディケータとしては効果的でないと考える．また，全体的に S/MIME で使用されている署名マークは視認性が低く，インディケータが持つ意味を正確にメール受信者に伝えらていないことが分かった．署名マークは，メール受信者が確認する情報である差出人欄や本文から離れた場所にあり，マークも小さい．そして，安全を示すマークと危険を示すマークが明確に区別されていないことが問題であると考える．図 4 は安全を示す署名マーク，図 5 は危険を示す署名マークであるが，両方の画像を見比べても明確に区別されているとは言い難い．メールは 1 日に何通も確認することがあるが，その行為の中で署名マークを識別することは，認知的負荷が大きいと言える．

また，ストレスの多寡による効果的なインディケータへの影響については，ストレスに関する回答の平均点が 2.5 より大きい回答者を「ストレスが多いユーザ群」，合計得点の平均点が 2.5 以下の回答者を「ストレスが少ないユーザ群」とした．それぞれのユーザ群にとって効果的なインディケータは，回答者の割合による有効性の順位の変動はあったものの，大きな差は見られなかった．

図 4: 安全を示す署名マーク

図 5: 危険を示す署名マーク

以上より，なりすましメール対策として提案するインディケータは既存対策のインディケータと比較して，視認性等の面で優れているという結果が得られた．そして，ユーザ特性としてメールを確認する環境下でのストレスの多寡について調査し，効果的なインディケータへの影響も調査したが，大きな差異は見られなかった．今後は，実際のメール環境でインディケータを使用し，馴化や作業効率への影響等を調査する必要がある．また，ストレスの多寡による効果的なインディケータへの影響については，ユーザ群を 2 分化する閾値の変更やインディケータが持つ意味の誤認識等，別の観点からも調査することが必要であると考える．

4　今後の予定

今年度は標準化活動を通じ，あるいは標準化された技術の新たな利用方法を考えることにより，幅広いステークホルダーにセキュリティに関する知識を啓蒙していく研究を行った．来年度も継続し，CYBEX WG のみならず他の WG，あるいは全世界で創出される素晴らしいセキュリティ技術を，より多くのステークホルダーに広めていく活動を継続していきたい．

Draft new Recommendation ITU-T X.1212 (X.cogent)

Design considerations for improved end-user perception of trustworthiness indicators

1 Scope

A wide variety of attacks employ replicated content from trustworthy service providers, thereby deceiving end-users into believing their false trustworthiness. Recommendation ITU-T X.cogent describes design considerations for improved end-user perception of trustworthiness indicators. The appendices describe representative techniques for measuring the end-user perception of such indicators.

2 References

None.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 disability [b-ITU-T F.790]: This is defined as a state when use of telecommunications equipment and services is restricted. Mainly, "disability" is viewed as a result of temporary or permanent functional limitation due to disease, accident, ageing and so on. More generally, "disability" includes a state when full use of telecommunications equipment and services is not possible due to the physical and/or social environment (e.g., voice telephony under noisy environment).

3.1.2 measurement [b-ENISA]: The act or the process of measuring, where the value of a quantitative variable in comparison to a (standard) unit of measurement is determined.

3.1.3 metric [b-ENISA]: A system of related measuring enabling quantification of some characteristic of a system, component or process. A metric is composed of two or more measures.

3.1.4 personally identifiable information (PII) [b-ITU-T X.1252]: Any information a) that identifies or can be used to identify, contact, or locate the person to whom such information pertains; b) from which identification or contact information of an individual person can be derived; or c) that is or can be linked to a natural person directly or indirectly.

3.1.5 phishing [b-ITU X.1254]: A scam by which an email or web user is duped into revealing personal or confidential information which the scammer can then use illicitly.

3.1.6 telecommunications accessibility [b-ITU-T F.790]: For the telecommunications area, the usability of a product, service, environment or facility by the widest possible range of users and especially users with disabilities.

3.1.7 person with disabilities [b-ITU-T F.791]:The correct way to refer a person with a disability [b-UNCRPD]..

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 trustworthiness indicators: Symbols presented by a web user agent that will be used to inform the trustworthiness of the website to end users.

4 Abbreviations and acronyms

DKIM DomainKeys Identified Mail

DOM Document Object Model

FNE Fear of Negative Evaluation

SSL Secure Socket Layer

URL Uniform Resource Locator

5 Conventions

None.

6 End-user perception of trustworthiness indicators

Protocols for cybersecurity information exchange, as identified in the *Overview of cybersecurity information exchange* [b-ITU-T X.1500], may convey useful information for trustworthiness decisions of any interactions in the cyberspace. Such information includes, but is not limited to, Extended Validation certificate information [b-CAB-Baseline], Level of Assurance of identities [b-ITU-T X.1254], DomainKeys Identified Mail (DKIM) signatures of e-mail [b-IETF RFC6376], and indication of phishing sites [b-IETF RFC5901].

These trustworthiness indicators are however often ignored or least considered by end users, according to past studies based on diverse demographics (details are provided in Appendix II). Thus it is necessary to improve the end-user perception of trustworthiness indicators.

7 Techniques for improved end-user perception of trustworthiness indicators

In this clause, several techniques for improving end-user perception of trustworthiness indicators are presented. These techniques can be used individually or in combination, as desired or appropriate, to present trustworthiness indicators in a more recognizable manner.

7.1 Visual elements

Developers of trustworthiness indicators shall consider the use of standardized visual elements. Past studies have revealed that symbolic encoding of trustworthiness indicators, e.g., in Uniform Resource Locators, are not friendly to novice users and they are often ignored [b-Miyamoto]. It is thus recommended to introduce visual elements, e.g., icons that represent trustworthiness indication. Implementers may consider employing a few standardized visual elements, as in road signs, to minimize cognitive overhead and training overhead.

According to product safety signs and labels [b-ANSI-Z535.4**]**, the use of signal words (e.g., “Danger,” “Warning,”) with associated colours (red, orange, yellow) decreases levels of risks.

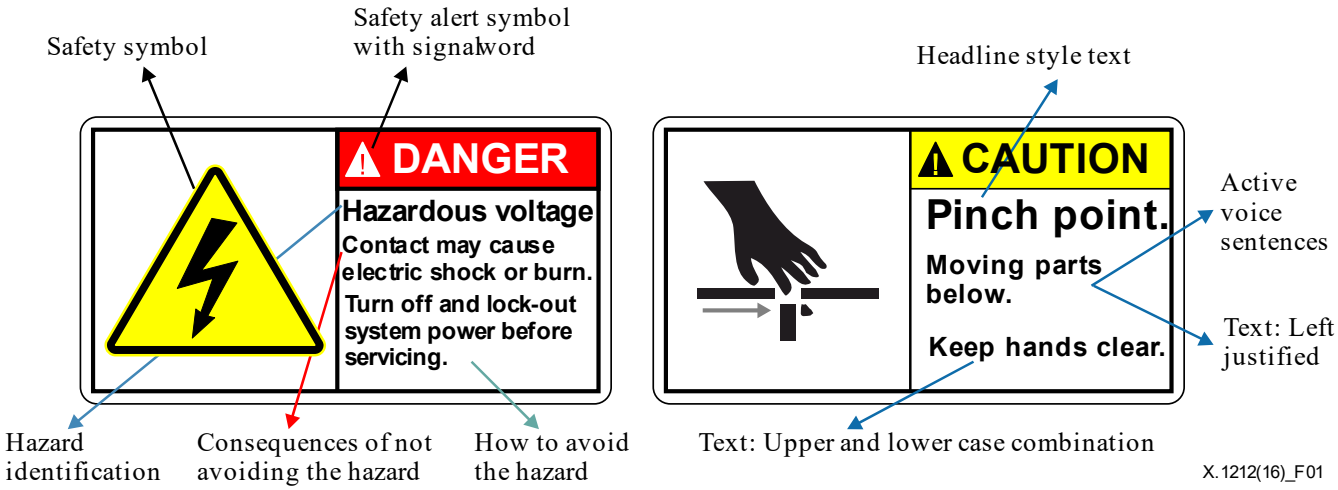


Figure 1 –Product safety signs and labels (ANSI Z535.4)

The message "DANGER" uses a white triangle, red exclamation mark and red background. The "WARNING" message employs a black triangle with an orange exclamation mark. The "CAUTION" message uses a black triangle with a yellow exclamation mark.

Additionally, developers of trustworthiness indicators should employ standard colouring schemes to represent the level of trustworthiness. In the context of colour psychology, red is used for attention. Red is the longest wavelength in the visible light spectrum, and has the property of appearing to be nearer than it is. Red therefore grasps users' attention and is used for traffic lights. The wavelength of yellow is relatively long and essentially stimulating, and it can grab users' attention. The center of the spectrum is green; it is the intermediate wavelength of visible light. Green also tends to require no adjustment to see, so it is used as a restful and relaxed sign. Blue calms the mind and aids concentration.

Developers of trustworthiness indicators may use the concept of the "social brain," which encourages pro-social and cooperative behaviour. Past studies have found that people behave in a more socially conscious manner when they are near images of watching eyes [b-Rigdon, b-Senju]. However, there is a sceptical view about it, which claims that the image of watching eyes had little to no effect on behaviour [b-Felt2014].

7.2 Narrative elements

Past studies have revealed that certain groups of users make their trustworthiness decision based on narrative writings, rather than domain names, protocol types or uniform resource locators (URLs) [b-Felst2014, b-Felt2015]. It is recommended to equip end-user software with the capability to convert symbolic information into narrative elements that do not employ acronyms. It can also be helpful to visually impaired users, when combined with text-to-speech systems.

In order to capture users' attention, i.e., warning messages, end-user software may need to consider several design criteria as follows:

1. Developers of trustworthiness indicators should avoid using technical terms. In the warning message, technical terms should be replaced with phrases or expressions that can be understood by users; they will ignore the message if they do not know how to properly respond to it.
2. Developers of trustworthiness indicators should consider the brevity of messages. Large quantities of text will indicate much effort to read, thus users may not read it. In the message, redundant text should be removed in order to be concise as well as being accurate.

It should be noted that there is a trade-off relationship between brevity and accuracy; it is not possible to explain all aspects of the threat model in a single short paragraph. Therefore warnings may utilize both visual and text elements. In order to calculate the level of the brevity, the developers may employ a readability index, which is the measure of readability that estimates the years of education a person needs to understand a piece of writing.

3. Developers of trustworthiness indicators should describe the risk that had occurred or is about to occur. Warning messages should describe the underlying risk, since users are likely to comprehend and comply with the message if it describes the risks explicitly and unambiguously. The message should also include instructions on how to avoid risk, unless these instructions are obvious in the statement of the risk.

7.3 Peripheral design transitions

Developers of trustworthiness indicators may test their interface regarding to the peripheral design transitions. Sudden transition in peripheral vision may be effective to signal potential risk. It is thus recommended to employ this technique through the transition of peripheral designs (typically called "themes" or "skins"), whenever end-users are faced with high-risk websites or e-mail messages.

7.4 Training mode

Developers of trustworthiness indicators may prepare training modes. The end-user perception of risk will be inaccurate at best if he or she is very rarely exposed to such risks. It is therefore recommended to equip end-user software with a training mode, where emulated risk events can be artificially generated and the end-user's perception accuracy can be trained. Such training can also be incentivized by gamifying the training.

7.5 Accessibility

Developers of trustworthiness indicators should design its interface considering accessibility. Vision refers to the ability to distinguish the form, size, shape and colour of visual stimuli. For individuals with vision impairment, there can be difficulties to find trustworthiness indicators. Due to the effects known as "protanopia" and "deuteranopia," some end-users have problems in distinguishing colours, e.g., red from green.

ISO/IEC defined the accessibility guideline document [b-ISO/IEC-40500:2012] for persons with disabilities, although, it does not directly address trustworthiness indicators on the address bar. The CA Browser Forum's baseline requirements document [b-CAB-Baseline] defines the standard for certificates and certificate authorities, although it does not define how browsers present certificates to users.

The telecommunications accessibility checklist [b-ITU-T-FSTP-TACL] ensures that the specified services and features are usable by diverse users, including persons with disabilities. In order to provide better accessibility for visual impairment or blindness, the interface should provide media presentation to the user, and have the ability to be controlled in various modes and types of control action. For persons with cognitive disabilities, important points should be highlighted to draw their attention as well as using supplemental media, such as icons, video and audio.

Screen reader applications may retrieve trustworthiness indicators from websites. They may present security information, e.g., the green address bar of an EV-SSL certificate, and read the information with text-to-speech services. They may also summarize information from a document object model (DOM) tree within the browser.

7.6 Children

With regard also to children on line a parent normally checks up by listening or seeing to the proceedings or activities of their children communicating online or has the information to restrict access in accessible format. That "protective" route may not be accessible to a parent with disabilities. That specific role as identified falls between two areas - child protection on line and accessibility for an adult/parent with disabilities with responsibilities for the upbringing of children without disabilities as well as children with disabilities.

Appendix I

Considerations for cognitive task analysis in cybersecurity

(This appendix does not form an integral part of this Recommendation.)

I.1 Considerations for cognitive task analysis in cybersecurity

Cognitive task analysis for cybersecurity purposes may involve the measurement of behavioural elements as well as the analysis of interactions, ultimately leading to the inference of internal mental processes. This Recommendation considers the three concepts of the information security, namely confidentiality, integrity, and availability, as the requirements for cognitive task analysis in cybersecurity.

I.2 Three enabling concepts of information security

Confidentiality

Measured data may include personal information, which is essentially privacy sensitive. Thus, the use of such data needs to be conducted carefully, accompanied by agreement with end users. The extent of sharing such information must be under strict control.

Integrity

The measurement methods might make use of the collected information regardless of the Fear of Negative Evaluation (FNE). Observations are often affected by FNE, in which some of people will conceal their human errors, as disclosing mistakes often damage their own self-image and professional standing.

Availability

Observations should employ the method which is easily applicable to people. Within the context of phishing prevention, the methods should be available while users are browsing presented information. Non-contact devices will be preferred. Furthermore, users will not carry implants or other devices that may hurt them in any way.

I.3 Possible measurement methods

Research on experimental psychology has evidenced a strong link between eye movements and mental disorders [b-Crawford,b-Noris]. Leigh et al. [b-Leigh] classified the eye movements into four categories, namely saccades, fixations, smooth pursuit movements, and vestibulo-ocular reflexes. Generally, the saccadic eye movement changes with what a person is seeing. In the context of mental model, Irwin et al. showed that mental rotation is suppressed during the movements [b-Irwin], and Tokuda [b-Tokuda] showed that mental workload, the indicator of how mentally/cognitively busy a person is, can be estimated from saccadic intrusions.

Validation of facial skin temperature is also feasible to gather information as a physiological measure of mental status [b-Or,b-Wang,b-Volskamp]. According to Genno et al. [b-Genno], their experiments showed that there are temperature changes in nose area when subjects experienced sensations like stress and fatigue. Furthermore, the thermography, when combined with other modes of measurement, provides a highly automated and flexible means to objectively evaluate workload [b-Or].

Aside from these solutions, brain activity, skin conductivity, heart measure, and blood pressure are often used to gather information, however, they tend to require obtrusiveness for users. Recognition of facial expression and gestures are helpful with regard to availability, however, they are easily affected by FNE.

Appendix II

Consideration of end user protection with cognitive task analysis

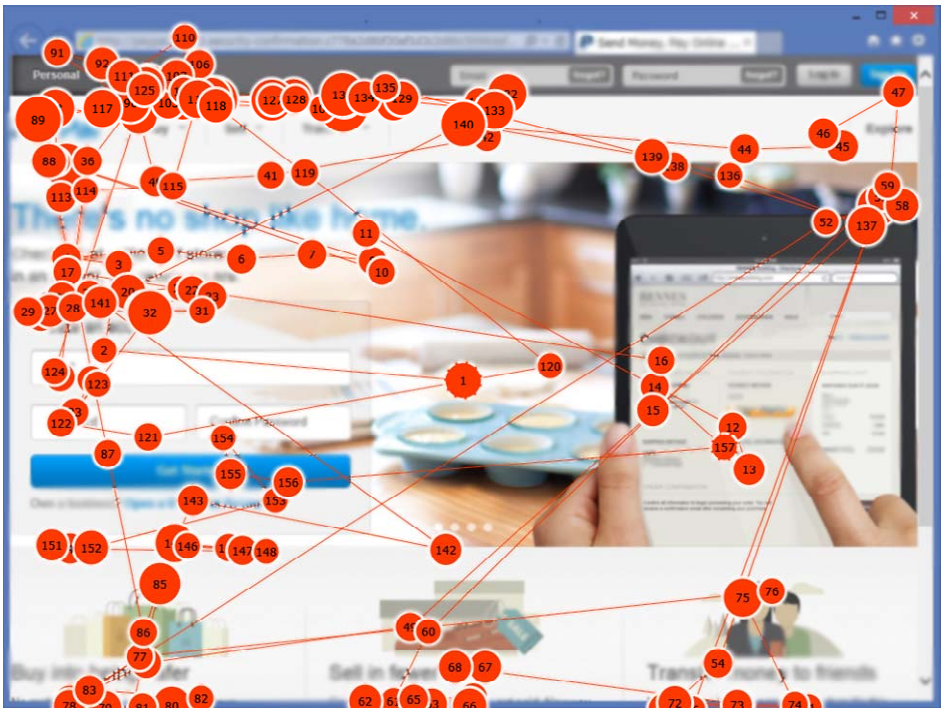
(This appendix does not form an integral part of this Recommendation.)

II.1 Estimation of users' knowledge and skills

A past study illustrates that end users can be categorized into two types, namely experts and novices [b-Miyamoto]. The experts evaluate a site's URL and/or browser's secure socket layer (SSL) indicator rather than the contents of a web page to judge the credibility of sites. On the other hand, novices received strong signals from web contents. Due to the nature of phishing, the web contents are quite similar to that of legitimate site, leading novices to fall victims to the phishing trap.

These distinct characteristics of end users are useful to adjust phishing prevention for each of them. A possible solution is to provide phishing detection with lower false negative for novices, and lower false positive for experts. Generally, phishing prevention systems have a problem in detection accuracy, because there is a trade-off relationship between false positive (labelling legitimate sites as phishing) and false negative (labelling phishing sites as legitimate). The false positive would increase if the systems focuses on decreasing false negatives (labelling phishing sites as legitimate). Reduction of both errors is considered difficult. In spite of that, the system must protect novices, who often fail to make the correct decision.

Using an eye-tracking device facilitates the identification of novices among web users. Figure II.1 shows the eye movement of a novice in a phishing website, and Figure II.2 shows that of an expert. Circles denote fixations, and the numbers in the circles denote the order of the fixation. In the phishing case, the novice looked at the web content but ignored the browser's address bar while assessing credibility, as shown in Figure II.1.

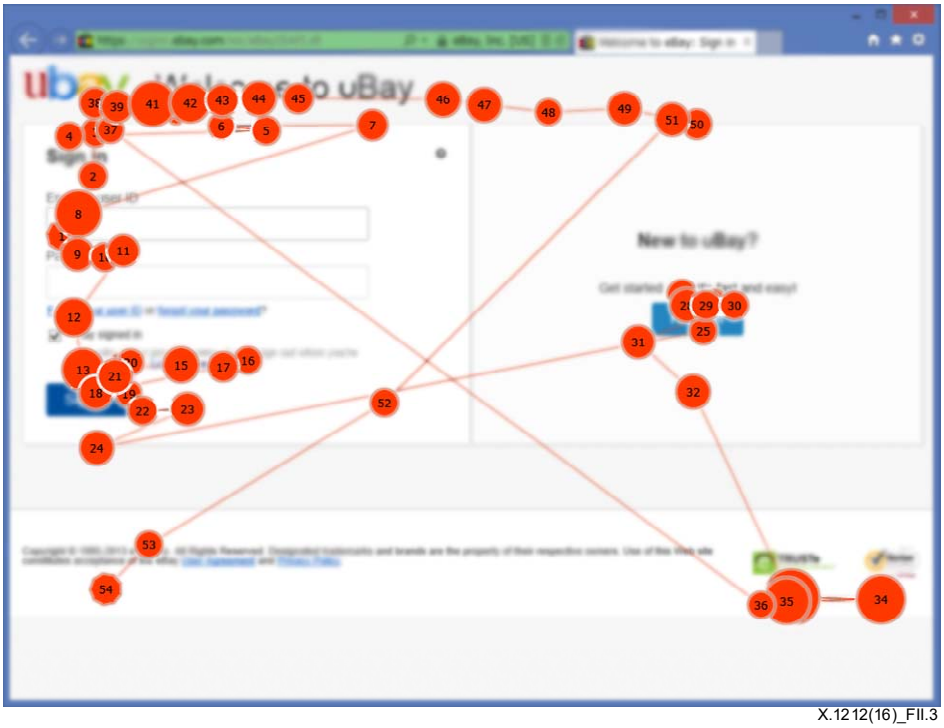


X.1212(16)_FII.1

Figure II.1 – A novice user on a phishing website

X.1212(16)_FII.2

Figure II.2 – An expert on a phishing website



X.1212(16)_FII.3

Figure II.3 – A novice on a legitimate website

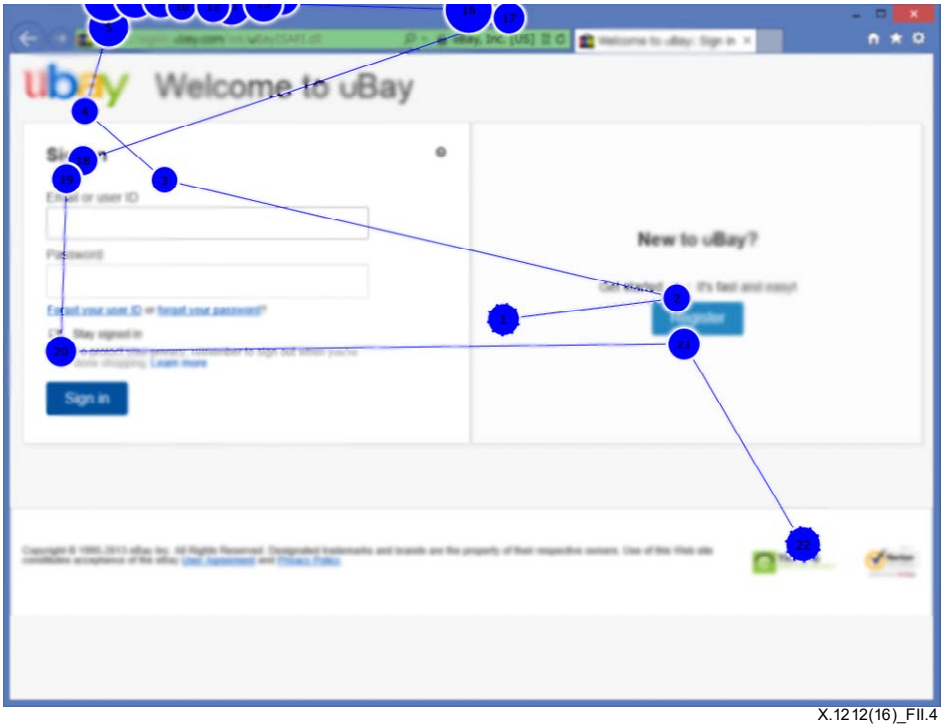


Figure II.4 – An expert on a legitimate website

In the legitimate case, the user also only paid attention to the web content as shown in Figure II.3. By contrast, an expert tends to evaluate the site's URL and/or the browser's SSL indicator rather than the contents of the web page in order to judge the credibility of the sites, as shown in Figure II.4. These behaviour observations indicate that experts tend to look at the address bar where the URL and browser's SSL indicator is displayed at the beginning of browsing. Novices are not aware of them due to the lack of knowledge on URL or SSL indicators.

Bibliography

[b-ANSI-Z535.4] ANSI (2011), Product Safety Signs and Labels.

[b-CAB-Baseline] CA/Browser Forum (2011), *Baseline Requirements for the Issuance and Management of Publicly-Trusted Certificates* Version 1.0. http://www.cabforum.org/Baseline_Requirements_V1.pdf

[b-ENISA] ENISA (V6_2, 2011), *Measurement Frameworks and Metrics for Resilient Networks and Services: technical report.*

[b-IETF RFC5901] IETF RFC 5901 (2010), *Extensions to the IODEF-Document Class for Reporting Phishing*. http://datatracker.ietf.org/doc/rfc5901/

[b-IETF RFC6376] IETF RFC 6376 (2011), *DomainKeys Identified Mail (DKIM) Signatures.* http://datatracker.ietf.org/doc/rfc6376/

[b-ISO/IEC-40500:2012] ISO/IEC 40500:2012, *Information Technology - W3C Web Content Accessibility Guidelines (WCAG) 2.0.*

[b-ITU-T-FSTP-TACL] ITU-T FSTP-TACL (2006), *Telecommunications Accessibility Checklist.* https://www.itu.int/publ/T-TUT-FSTP-2006-TACL

[b-ITU-T F.790] Recommendation ITU-T F.790 (2007), *Telecommunications accessibility guidelines for older persons and persons with disabilities.* https://www.itu.int/rec/T-REC-F.790

[b-ITU-T F.790] Recommendation ITU-T F.791 (2015), *Recommendation ITU-T F.791 (2015), Accessibility terms and definitions.* https://www.itu.int/rec/T-REC-F.791

[b-ITU-T X.1252] Recommendation ITU-T X.1252 (2010), *Baseline identity management terms and definitions.* https://www.itu.int/rec/T-REC-X.1252

[b-ITU-T X.1254] Recommendation ITU-T X.1254 (2012), *Entity authentication assurance framework.* https://www.itu.int/rec/T-REC-X.1254

[b-ITU-T X.1500] Recommendation ITU-T X.1500 (2011), *Overview of cybersecurity information exchange.* https://www.itu.int/rec/T-REC-X.1500

[b-Crawford] Crawford, T.J., Higham, S., Renvoize, T., Patel, J., Dale, M., Suriya, A., Tetley S. (2005), *Inhibitory control of saccadic eye movements and cognitive impairment in Alzheimer's disease*, Biological Psychiatry, vol. 9, no. 57.

[b-Felt2014] Felt, A.P., Reeder, R.W., Almuhimedi. H., Consolvo, S. (2014), *Experimenting At Scale With Google Chrome's SSL Warnings*, in Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems.

[b-Felt2015] Felt, A.P., Ainslie, A., Reeder, R.W., Consolvo, S., Thyagaraja, S., Bettes, A., Harris, H., Grimes, J. (2015), *Improving SSL Warnings: Comprehension and Adherence*, in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.

[b-Genno] Genno, H., Ishikawa, K., Kanbara, O., Kikumoto, M., Fujiwara, Y., Suzuki, R., Osumi, M. (1997), *Using facial skin temperature to objectively evaluate sensations*, International Journal of Industrial Ergonomics, vol. 19.

[b-Irwin] Irwin, D.E., Brockmole, J.R. (2000), *Mental rotation is suppressed during saccadic eye movements*, Psychonomic Bulletin and Review, vol. 7, no. 4.

[b-Leigh] Leigh, R.J., Zee, D.S. (1991), *The Neurology of Eye Movements*, 4th ed. Oxford University Press.

[b-Miyamoto] Miyamoto, D., Iimura, T., Tazaki, H., Blanc, G., Kadobayashi, Y. (2014), *EyeBit: Eye-Tracking Approach for Enforcing Phishing Prevention Habits*, in Proceedings of the 3rd International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security.

[b-Noris] Noris, B. Benmachiche, K., Meynet, J., Thiran, J.P., Billard, A. (2007), *Analysis of Head-Mounted Wireless Camera Videos for Early Diagnosis of Autism*, Advances in Soft Computing, vol. 45.

[b-Tokuda] Tokuda, S., Obinata G., Palmer, E., Chaparro, A. (2011), *Estimation of mental workload using saccadic eye movements in a free-viewing task*, in Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

[b-Or] Or, C.K.L., Duffy, V.G. (2007), *Development of a facial skin temperature-based methodology for nonintrusive mental workload measurement*, Occupational Ergonomics, vol. 7.

[b-Rigdon] Rigdon, M., Ishii, K., Watabe, M., and Kitayama, S. (2009), *Minimal social cues in the dictator game*, Journal of Economic Psychology vol. 30, iss. 3.

[b-Senju] Senju, A., Johnson, M.H. (2009), *The eye contact effect: mechanisms and development*, Trend in Cognitive Science.

[b-UNCRPD] United Nations, Conventions on the Rights of Persons with Disabilities (2006).

[b-Volskamp] Voskamp, J., Urban, B. (2009), *Measuring Cognitive Workload in Non-military Scenarios Criteria for Sensor Technologies*, in Proceedings of the 5th International Conference on Foundations of Augmented Cognition.

[b-Wang] Wang, L., Duffy V.G., Du, Y. (2007), *A composite measure for the evaluation of mental workload*, in Proceedings of the 1st International Conference on Digital Human Modelling.